Received: 23 March 2020 | Revised: 2 November 2021 | Accepted: 14 December 2021

DOI: 10.1002/hec.4468

RESEARCH ARTICLE

Health Economics WILEY

Moving out but not for the better: Health consequences of interprovincial rural-urban migration in China

Xinjie Shi[1,2,3]

[1]China Academy for Rural Development, School of Public Affairs, Zhejiang University, Hangzhou, China

[2]Australian Centre on China in the World & Crawford School of Public Policy, The Australian National University, Canberra, Australia

[3]Institute for Common Prosperity and Development, Zhejiang University, Hangzhou, China

Correspondence
Xinjie Shi, China Academy for Rural Development, School of Public Affairs, Zhejiang University, Hangzhou, China.
Australian Centre on China in the World & Crawford School of Public Policy, The Australian National University, Canberra, Australia.
Institute for Common Prosperity and Development, Zhejiang University, Hangzhou, China.
Email: xjshi1990@zju.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 72003170
The National Social Science Fund of China, Grant/Award Number: 21&ZD091
Zhejiang University-International Food Policy Research Institute Center for International Development Studies, Grant/Award Number: 126000-541902/006

Abstract

Using panel data from a survey conducted by the Research Center for Rural Economy, this study overcomes the selection bias associated with most health status measures by incorporating dialect proximity as an instrumental variable to evaluate the causal effects of interprovincial migration on individuals' health status in China. The results indicate that, compared to an absorbed category, which includes intraprovincial migration and non-migration, interprovincial migration worsens health (measured by the self-reported health status score) by 3%. It is also found that, compared to intraprovincial migrants, interprovincial migrants are more likely to report a lower SHS score, whereas the health effect differences between interprovincial migrants and non-migrants are insignificant. The study identifies two potential mechanisms that explain this result: (i) The increase in income following migration may improve interprovincial migrants' health, and (ii) the exposure to hazardous working environments worsens health to a significant extent. This finding is also linked to the following gender-related finding: Compared to women, men are more likely to be employed in heavy industries with hazardous working environments; the negative effects of interprovincial migration on the health status are stronger for men than for women migrants, suggesting that the industry effect dominates.

KEYWORDS

China, health, interprovincial migration, selection bias

1 | INTRODUCTION

Rural–urban migration—the internal outflow of rural labor to urban areas due to favorable social and economic opportunities—plays a pivotal role in determining the lives of millions of individuals in China. By the end of 2008, China had a total of 140 million rural–urban migrant workers, officially defined as those who have a rural *hukou* and work in the non-farm sector in a local county (local migrants) or outside the county (outgoing migrants)[1] for at least six months in the same year. In 2019, the number of migrants nearly doubled to 291 million, comprising 117 million local and 174 million outgoing migrants; furthermore, 43.1 percent of the outgoing migrants migrated outside the province. Literature shows that interprovincial migration played a pivotal role in wealth and population redistribution (Wu et al., 2018). In China, rapid development coupled with massive rural–urban migration significantly affects the health of the population in both rural areas—where the families of migrants are left behind—and urban areas—where migrants relocate in search of opportunities. Although the literature on the migration-associated health implications for the families left behind is rapidly expanding,[2] only a few studies empirically evaluate the impact of migration on the health status of migrating individuals. This study addresses the aforementioned research gap by comparing the health outcomes of interprovincial migrants with those of an absorbed category comprising intraprovincial migrants and non-migrants in China.[3]

Interprovincial migration can affect the health status of migrants through several channels. On the one hand, individuals' health status may worsen as a result of migration. At least two types of health issues are experienced by many migrants. First, compared to urban registered residents and individuals who stay in rural areas, migrant workers have a significantly higher incidence of workplace injury and occupational health problems in China. Since many migrants work in stressful and hazardous working environments, such as mining, manufacturing, and construction, they experience significant numbers of work-related injuries and fatalities. Moreover, five of the world's most fatal industrial accidents in the past decade occurred in China (Chan & Griffiths, 2010). In general, rural–urban migrants work long hours and experience high work intensity, which increases their health risks from worksite hazard exposure (Mou et al., 2013; Walsh & Walsh, 1987).

The second type of health issues faced by migrants is mental health issues arising mainly from the stress experienced by them in adapting to an unfamiliar community. This risk of mental disorders is exacerbated by the reduction of social support caused by migrants' separation from their families (Bhugra & Jones, 2001). They may experience some other mental problems caused or aggravated by stress arising from economic pressure (Song & Sun, 2016), stigmatization (Li et al., 2007), the discrepancy between expectations and reality (Zhang et al., 2009), and discrimination and difficulty in acculturation (Chen, 2011; Gui et al., 2012). These factors can affect not only the current psychological well-being but also future trajectories of mental health of the migrants (Mou et al., 2013). The majority of evidence indicates that rural–urban migrants experience severe psychological risks, among which depressive symptoms (Qiu et al., 2011) and insomnia are the most common.

Furthermore, several studies establish that, compared to intraprovincial (relatively short-distance) migrants, interprovincial (relatively long-distance) migrants are more likely to work in energy-consuming industries and experience mental health issues (Fan, 2005; Gries et al., 2016; Poncet, 2006; Su et al., 2018; Wu et al., 2018). One reason is that most interprovincial migration destinations offer higher salaries but at a relatively higher cost to health, such as greater work stress and longer working hours. Further, intraprovincial migration is characterized by migration over short distances, that is, intraprovincial migrants are more likely to visit their rural places of origin regularly (Su et al., 2018), rendering them less susceptible to the negative effects of family separation compared to interprovincial migrants.

However, changes in the degree and direction of such relationships over time remain uncertain (Pyakuryal et al., 2011). On the other hand, there is a major upside to rural–urban migration regarding the health status of migrants: they are likely to earn more by working in cities. Farmers typically migrate to urban areas to earn high incomes and improve their living standards (Gupta et al., 2009; Howell, 2017; Knight & Song, 2003; Knight et al., 1999; Scharf & Rahut, 2014). As shown in Figure 1, according to the Chinese Migrant Annual Report,[4] the monthly income of migrants steadily increased from 1340 yuan in 2008 to 3962 yuan in 2019. In addition, compared to intraprovincial migration, interprovincial migration was associated with a higher income, particularly among western Chinese migrants (Wu et al., 2018). When migrants earn high nonfarm incomes, they are more likely to have better living standards and healthcare facilities and invest more in such facilities (Lu, 2010; Song & Sun, 2016; Sun et al., 2014). In addition, interprovincial migrants may experience altered dietary patterns. This can be attributed to their increased income and the high spread of nutritional knowledge in large cities. However, the migrants' eating habits may be negatively affected by the long working hours and highly stressful work environments associated with urban workplaces; this impact may be more visible among the interprovincial migrants originating from a different dietary culture.

These contradictory theoretical predictions explain why studies on the health status of migrants in China present mixed results. Some studies reveal that rural–urban migrants in China experience various stresses arising from work-related, familial, and interpersonal difficulties and, hence, have a high incidence of health problems (Li et al., 2009; Wong et al., 2008). Utilizing the data from a household survey conducted in Beijing, Chen (2011) found that migration has a detrimental effect on the physical health status of migrants. In contrast, Song and Sun (2016) examined the impact of rural–urban migration on migrants' health status by using the panel data from 2003 to 2006 constructed by the Research Center for Rural Economy (RCRE) at the Ministry of Agriculture, China, and found that the impact of short-term migration on self-reported health was significantly positive, although long-term continuous migration had an insignificant impact (close to zero) on migrants' health status. This positive impact of short-term migration is explained mainly by the high

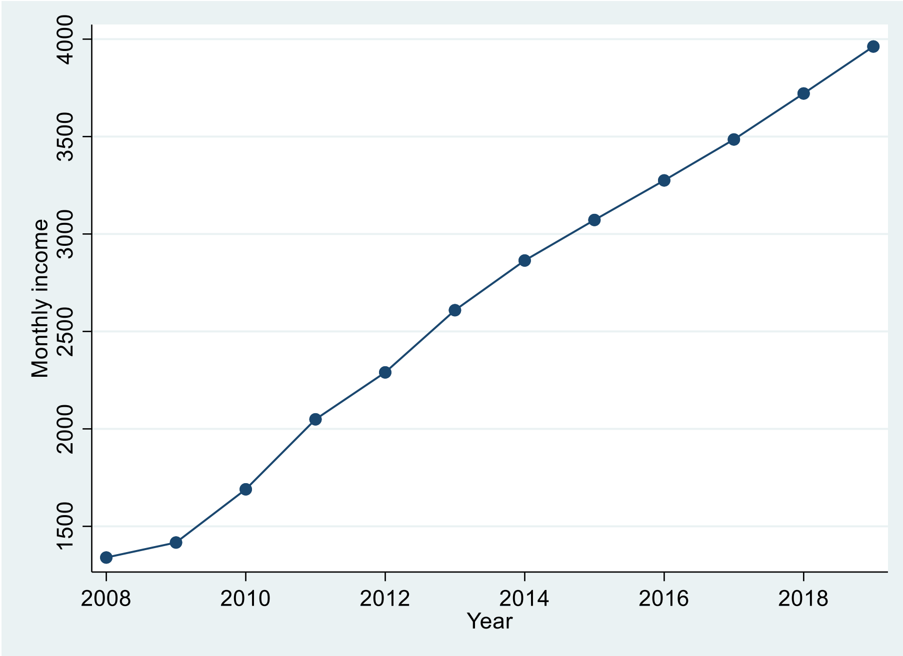

FIGURE 1 Migrants' monthly income for 2008–2019. *Source*: Chinese Migrant Annual Report

migration income; however, exposure to hazardous working environments and separation from family have no effect on the migrants' health status. Moreover, although it is important to distinguish between the health impacts of interprovincial and intraprovincial migration, little is known at present about their differential effects on individuals' health status.

A key challenge to credibly evaluating the impact of interprovincial migration on health is that migrants form a select group. Many studies examine the selection bias associated with the evaluation of rural–urban migration, termed the "healthy migrant hypothesis" by demographers: It states that rural–urban migrants are healthier than non-migrants on average. This can be further extended to the "healthy interprovincial migrant hypothesis." Some studies indicate that individuals with good health (e.g., young individuals) are more likely to migrate than their less healthy counterparts (De Brauw et al., 2002; Shen, 2013). However, interprovincial migrants may also need to be healthier because migration costs increase with an increase in the distance between the origin and destination locations (Poncet, 2006). Moreover, migrants with severe health problems tend to return to their rural origins (Biao, 2007; Hu et al., 2008), and the majority of interprovincial migrants are temporary migrants who ultimately return to their villages. Hence, this should be accounted for when estimating the causal impact of interprovincial migration on migrants' health status.

In contrast to Chen (2011) and Song and Sun (2016), who examined the impact of rural–urban migration on health, this study focuses on the impact of interprovincial migration and uses an innovative strategy to address the selection bias. It is the first to examine the causal effects of interprovincial migration on migrants' health status in China using an instrumental variable (IV) approach by adopting the dialect proximity between the migrants' source village and the province in which this village is located as an instrument for interprovincial migration. Based on a dataset constructed by the RCRE, the results of the IV strategy reveal that interprovincial migration worsens migrants' health. Notably, separation from family and changes in dietary patterns are not identified as mechanisms through which interprovincial migration affects health. Rather, the two main mechanisms are (i) an income effect, which makes interprovincial migration beneficial to health, and (ii) an industry effect, which is predominantly negatively associated with health due to migrants' exposure to hazardous working environments. Furthermore, compared to women, men are more likely to work in stressful working environments, and thus, the negative effects of interprovincial migration on the health status are stronger for men than for women. The negative impact of interprovincial migration is also smaller for those with strong assimilation compared to those with weak assimilation. A potential reason is that those with strong assimilation are less likely to work in heavy industries due to the availability of more job information via urban contacts. The main findings are reinforced through a series of robustness checks.

The remainder of this paper is organized as follows: Section 2 presents the main empirical strategy. Section 3 introduces the data used by the study and presents descriptive statistics. Section 4 presents the key findings on how interprovincial migration affects migrants' health status. Section 5 examines the main mechanisms of the causal effect and particularly emphasizes the role of gender. Subsequently, Section 6 presents a series of robustness checks that reinforce the main findings. Finally, Section 7 concludes the paper. In addition, the online appendix reviews the literature on identifying the causal impact of migration on health status and presents the IV identification strategy used in this study.

2 | EMPIRICAL STRATEGY

Since the study's primary goal is to estimate the effect of interprovincial migration on health status, the simplest econometric framework should first estimate the following equation to determine individuals' health status as a function of their migration status and other demographic and economic characteristics:

$$y_{it} = \alpha + \beta M_{it} + \gamma X_{it} + \varepsilon_{it} \tag{1}$$

where the dependent variable y_{it} denotes the health status of individual i at time t. Further, the effect of interest is captured by the coefficient of M_{it}, a dummy variable that indicates whether individuals migrate to other provinces. M_{it} is equal to 1 for interprovincial migrants (those who migrate relatively long distances) and 0 for an absorbed category of intraprovincial migrants (those who migrate relatively short distances) or non-migrants. The vector of covariates, X_{it}, incorporates the following individual characteristics: gender, age, square of age, education (years of schooling), two dummy variables associated with farm and non-farm training, and a variable denoting the average health status of household members excluding the respondent. To better understand the mechanisms by which migration affects health status, I utilize respondent information pertaining to income, industry type, separation status, and dietary pattern.

One concern regarding Equation (1) is that (pooled) ordinary least squares (OLS) estimation generally provides a biased estimate of β because the migration indicator is likely to be endogenous. This may be a major issue in identifying a causal link if there is a strong correlation between unobserved components that influence both migration choice and migrants' health status. Moreover, individuals with better health status are more likely to migrate relatively longer distances than others, which creates a reverse causality issue. The first identification strategy to address this issue relies on an individual fixed-effects model that links changes in migration status with changes in health status:

$$y_{it} = \alpha + \beta M_{it} + \gamma X_{it} + \varnothing_i + \varepsilon_{it} \tag{2}$$

The main difference between Equations (1) and (2) is that Equation (2) includes an individual-specific effect, $\varnothing_i$, that controls for all observed and unobserved time-invariant factors.[5] This empirical strategy addresses concerns regarding time-invariant endogeneity sources. However, it does not address time-varying sources of endogeneity in some cases, for example, when a shock to household income compels an individual to migrate and, thereby, changes their health status. Moreover, this strategy cannot account for the "healthy interprovincial migrant hypothesis."

To further evaluate the extent of a causal relationship, I use a second identification strategy, which considers a valid IV, as discussed in the Appendix. As shown by Antman (2011), time-varying instruments are ideally coupled with individual fixed effects to produce a fixed effect–instrumental variable (FE-IV) estimator. However, in practice, it is challenging to find such instruments that both are valid and vary with time. The instrument considered in this study is a dummy indicating whether a village's dialect is the same as the dominant provincial dialect, that is, it is a time-invariant instrument. Hence, the main empirical model enables the estimation of Equation (1) using the IV approach, in which migration status is estimated using the following first-stage regression:

$$M_i = \rho Z_i + \sigma X_i + \epsilon_i \tag{3}$$

where Z is the dialect instrument excluded in Equation (2). This calculation yields an OLS-IV estimator. If the instrument is valid, the OLS-IV estimator is consistent, as demonstrated by Antman (2013a). The main threat to this identification strategy is the exclusion restriction necessary for instrument estimation. For instance, dialect proximity may be associated with professional health services and, thereby, influence migrants' health services, rendering it invalid as an instrument variable. This consideration violates the exclusion restriction of the IV identification strategy that is necessary for instrument estimation. However, I discuss that this is not the case in Section 4.2.

3 | DATA AND DESCRIPTIVE EVIDENCE

3.1 | Data

The main dataset used in this study was obtained from a survey conducted by the RCRE. The center has been regularly conducting this survey since 1986 (except in 1992 and 1994), when the initial sample, which was checked annually, was selected using stratified random sampling and a panel dataset with a small exit and entry rate was created. The RCRE survey is the largest longitudinal household-based survey conducted in rural China and covers more than 20,000 households in approximately 300 villages in 31 provinces. The survey data on the average rural income and expenditure are comparable to those published in the official Chinese Yearbook of Statistics.

The dataset contains various individual- and household-level statistics, including income, consumption, demographic information, and agricultural production information. In this study, I used the data from 2011 to 2014; I excluded observations with missing information to obtain a sample of 172,461 individuals for the 4 year period. In addition, I used data provided in the *Language Atlas of China* (Wurm et al., 1987) to classify the sampled village dialects. The Chinese dialectal family tree depicted in Figure S2 reveals that Chinese dialects can be classified into 10 supergroups and further sub-divided into 20 groups and 105 subgroups (Liu & Xu, 2016). As the dialects can be identified only at the county level, I identified the county in which each village is located and assigned the county's dialect to the corresponding village. I used these data to construct the aforementioned instrument; this is discussed further in Section 4.

3.2 | Key variables and descriptive statistics

Based on whether a certain household member left the village, the duration that he or she stayed in cities in the past year, and the place to which this person migrated, I defined migrants as individuals with rural *hukou* who stayed in urban areas for more than 6 months of the same year. This is the official definition of a migrant as provided by the Chinese government[6] and is widely used in the literature (Lu et al., 2017). Table 1 reveals that 12 percent of migrants chose to migrate to other provinces (long-distance migration). I further divided the complete sample into two groups, interprovincial migrants and others (including within-provincial migrants and non-migrants), and provided summary statistics for the key variables for the complete sample and separately for each group.

TABLE 1 Descriptive statistics of the key variables for the pooled sample

	Full sample		Intraprovincial migrants and non-migrants		Interprovincial migrants	
Variable	**Mean**	**Std.Dev.**	**Mean**	**Std.Dev.**	**Mean**	**Std.Dev.**
Health (SHS)	4.57	0.72	4.55	0.73	4.71	0.56
Medical expenditure	1.91	4.18	1.88	4.20	2.09	4.06
Inter-provincial migration	0.12	0.32	0.00	0.00	1.00	0.00
Gender	0.51	0.50	0.49	0.50	0.64	0.48
Age	39.64	12.94	40.44	13.06	33.68	10.20
Education	7.72	3.08	7.64	3.11	8.35	2.69
Income (log, yuan)	9.42	0.68	9.42	0.69	9.41	0.58
Non-farm training	0.05	0.22	0.05	0.21	0.08	0.28
Farm training	0.04	0.19	0.04	0.19	0.03	0.16
Family health	4.58	0.60	4.58	0.60	4.53	0.61
Industry	0.39	0.49	0.37	0.48	0.55	0.50
Separation	0.74	0.44	0.74	0.44	0.75	0.43
Food ratio	3.31	5.30	3.36	5.51	2.99	3.26
Observations	172,461		128,195		44,266	

Note: Income was calculated as the logarithm of per capita household income. Medical expenditure was calculated as the logarithm of per capita household medical expenditure. Family health indicates the average health status of household members, excluding respondents.

Abbreviation: SHS, self-reported health status.

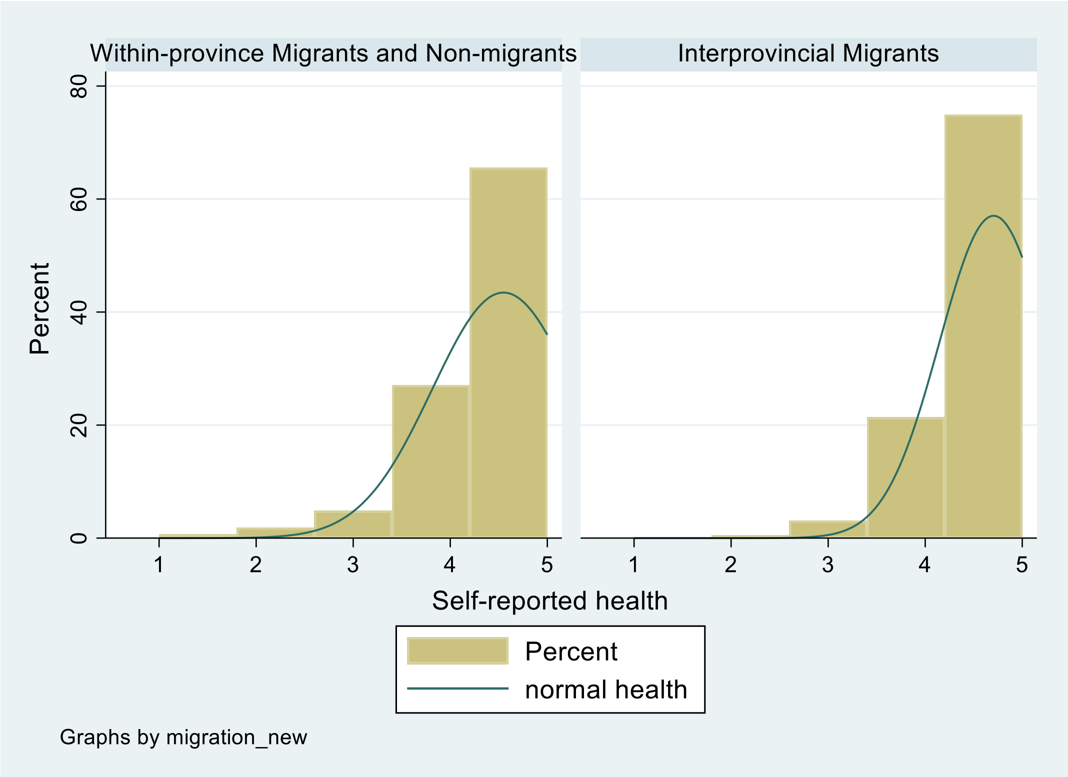

FIGURE 2 Distribution of self-reported health status in the study sample

In this study, health status was the dependent variable. I used self-reported health status (SHS) as the main measurement based on the following statement and recoded it into a new variable with acceptable values ranging from 1, very poor, to 5, excellent: "Your current state of health: 1. Excellent 2. Good 3. Average 4. Poor 5. Very Poor."

SHS was selected as the proxy for health because it not only is the only relevant health measure considered in the survey but also has been widely used in the literature (Antman, 2010; Ao et al., 2016; Currie & Stabile, 2003; Song & Sun, 2016). There are at least twofold benefits to using SHS, as discussed by Baker et al. (2004): "The first is greater comparability of the measures across respondents; the second is a lower likelihood that respondents rationalize their own behavior through their answers." In addition, since the questionnaires were collected immediately before the Chinese New Year, when the majority of the migrants returned to their hometown to celebrate the festival with their family, the migrants themselves were able to report their health status. This enhanced the reliability of the health data (Song & Sun, 2016).

Although the dataset used in this study matches the national means of important variables, a limitation of this study is that it uses a single health status variable.[7] Furthermore, self-reported health status measures may be susceptible to bias (Baker et al., 2004). The main concern regarding SHS is that it is inherently subjective and may thus be associated with personal characteristics. In this respect, the fixed-effects strategy used in this study addresses any bias associated with any observed or unobserved individual trait, because it considers variations occurring within the same individual over time (Antman, 2013a). Furthermore, despite being susceptible to several biases, self-reported health status measures effectively predict other more objective health indicators, such as mortality (Idler & Benyamini, 1997; Kuhn et al., 2006), fitness (Hertzman et al., 2001), morbidity (Power et al., 1991), quality-adjusted life years (Cutler et al., 1997), disability-adjusted life years (Musgrove, 1993), and the quality of well-being scale (Kaplan & Anderson, 1988).

Table 1 shows that the mean value of the health status is slightly higher for interprovincial migrants compared to others. Figure 2 reveals that most individuals report having good health, and interprovincial migrants have slightly better health compared to within-provincial migrants and non-migrants. However, it cannot be concluded that interprovincial migration positively affects migrants' health status as a result of the aforementioned selection bias.

Further, Table 1 presents the summary statistics for other variables, including gender, age, years of schooling, income, medical expenditure, non-farm training, and farm training. I controlled for the average health status of household members, excluding respondents, as well. Moreover, to represent the employed industry outcomes, I used a dummy variable set to one when the industry's workplaces included dangerous working environments, such as mining, manufacturing, construction, and transportation and 0 otherwise. Additionally, the separation variable was a dummy with a value of one if an individual was separated from other family members and 0 otherwise. In this study, food ratio was considered a proxy for diet; it was defined as the ratio of expenditures on non-staple food and staple food. Table 1 reveals that most interprovincial migrants are young men; they are more likely to attain higher education, spend more on medical issues, and attend non-farm training. Moreover, interprovincial migrants are more likely to be employed in heavy industries and, thereby, be separated from other family members.

4 | RESULTS

4.1 | Main results: Interprovincial migration and migrant's health status

Table 2 depicts the results of regression analysis. Column 1 presents the baseline OLS results based on Equation (3) and indicates that the coefficient of migration is 0.056. This suggests that rural–urban migration is positively associated with interprovincial migrants' health status. The results for other variables are as expected. For example, the impact of age on the health status is expressed using an inverse U-shaped curve; furthermore, men with good education and healthy family members tend to have good health status. Moreover, I conducted a fixed-effect regression based on Equation (4), whose results are reported in Column 2 of Table 2. These results reveal a significant decrease in the coefficient of migration

TABLE 2 Effects of interprovincial migration on migrants' health status

	(1) OLS	(2) FE	(3) IV	(4) IV	(5) IV	(6) IV
Inter-provincial migration	0.056***	0.034***	−0.138***	−0.512***	−0.271***	−0.071
	(0.01)	(0.01)	(0.05)	(0.13)	(0.04)	(0.05)
Gender	0.091***		0.101***	0.162***	0.051***	0.098***
	(0.01)		(0.00)	(0.01)	(0.00)	(0.00)
Age	0.015***	0.024***	0.017***	0.003	0.010***	0.016***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Age squared	−0.036***	−0.051***	−0.040***	−0.023***	−0.032***	−0.038***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Education	0.019***	0.006***	0.019***	0.022***	0.003**	0.021***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Non-farm training	0.037**	0.031***	0.047***	0.041***	0.033***	0.044***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Farm training	0.066***	0.006	0.061***	0.077***	0.033**	0.061***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Family health	0.580***	0.386***	0.576***	0.678***	0.494***	0.588***
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
Constant	1.743***	2.671***	1.757***	1.537***	2.592***	1.697***
	(0.06)	(0.06)	(0.02)	(0.19)	(0.05)	(0.02)
N	172,461	172,461	172,461	90,920	43,982	145,086
Adjusted *R*-squared	0.342	0.825	0.335	0.298	0.297	0.334
F	2128.266	564.901	6254.647	3028.954	1217.733	5526.614

Note: ***1%,**5%, *10%; robust standard errors are provided within parentheses. The dependent variables are the SHS for all the columns. In Columns 1–4, the independent variable of interest is an indicator for interprovincial migration that equals 1 for interprovincial migrants and 0 for an absorbed category of intraprovincial migrants and non-migrants. The results of Columns 1–2 are based on OLS and fixed-effect models, respectively; the result of Column 3 adopts dialect as an IV, and Column 4 further restricts the sample to those aged 40 and above. In Column 5, in which non-migrants are excluded, inter-provincial migration equals 1 for interprovincial migrants and 0 for intraprovincial migrants; in Column 6, in which intraprovincial migrants are excluded, interprovincial migration equals 1 for interprovincial migrants and 0 for non-migrants.

Abbreviations: FE, fixed effect; IV, instrumental variable; OLS, ordinary least squares.

(0.034) when individual fixed effects are controlled for. The magnitude and significance of other coefficients do not change significantly.

However, it is possible that both the OLS and fixed-effect estimates are biased. Specifically, health-related return migration may occur under the policy context of the New Rural Cooperative Medical Scheme (NRCMS). Additionally, the fixed-effects estimates do not consider time-varying health, which can cause migrants to return home to obtain medical assistance and insurance-covered health care in the absence of such health care in their destination cities, particularly those outside their own provinces. Hence, I used the IV strategy to better address this reverse causality issue due to the IV's rationality. The first stage, depicted in Table S2, reveals that the IV proximity of dialect significantly impacts interprovincial migration. The highly significant negative coefficient of the dialect variable indicates that individuals living in the village whose residents speak the dialect that is dominant in the province are less likely to migrate to other provinces. This implies that dialect influences individuals' interprovincial migration decisions.

The coefficient on migration, reported in Column 3 of Table 2, is the LATE of interprovincial migration estimated using the instrument strategy. If there is positive selection for migration, I expect this coefficient to be lower than the OLS estimate. This turns out to be the case, so much so that there is a sign reversal. This suggests that interprovincial rural–urban migration decreased the SHS score by 0.138. The average SHS score is 4.57, indicating that interprovincial migration decreases the SHS score by 3%. Notably, the IV results reverse the sign of the results using the OLS and fixed-effects model. This can be attributed to the healthy migrant hypothesis discussed earlier. The OLS and fixed-effects models indicate that rural residents are more likely to migrate long-distance (e.g., interprovincial migration). However, migration worsens migrants' health status,[8] which is identified by the IV results, rather than the OLS and fixed-effects models as neither of them successfully address the selection bias.

This result is consistent with the finding by Black et al. (2015) that the Great Migration—a landmark in US history—reduced the longevity of those who migrated from the South. Contrary to their study, which used the proximity of migrants' birthplace to railroads as an instrument for migration, I used dialect proximity, which is widely recognized as an important factor affecting migration, as the instrument. This is one of the cases in which the IV results reverse the signs of the coefficient of interest. Other examples include Frankel and Rose (2005), who find that the IV results reverse the impact of trade openness on rural water access. However, the robustness of IV results rely on the validity of the IV; the following section shows that dialect proximity is a valid IV in a number of aspects.

In addition, since intraprovincial migrants and non-migrants probably differ in many aspects, treating them as an absorbed category could be problematic. In Columns 5 and 6 of Table 2, I compare interprovincial migration with non-migrants and intraprovincial migration separately. Compared to intraprovincial migrants, interprovincial migrants are found more likely to report a lower SHS score by 0.271, whereas the health effect do not differ significantly between interprovincial migrants and non-migrants.[9]

4.2 | Issues associated with exclusion restriction

The results may be affected by challenges related to the high mobility of the present-day Chinese society and the fact that most young people learn Mandarin at school. Therefore, dialect proximity may not be a good predictor of interprovincial migration for this group. Hence, I restricted the sample to include only individuals aged 40 years and above and re-estimated the results using dialect proximity. The results depicted in Column 4 of Table 2 reveal that the negative effect of interprovincial migration on migrants' health status is significant (and large) for older people. This provides additional evidence to support the IV's validity and robustness of the main results.

The validity of the inferences is also determined by the assumption that the dialect instrument affects only the health status as a result of migration and has no other direct impact. One concern is that it may be easier to recruit medical professionals if people migrate to a region without experiencing language barriers. The best method to examine this possibility is to identify the language barriers existing in the migration destination. However, adequate data on the relevant language barriers was not available. An alternative was to investigate whether the village's dialect was correlated with local health professionals.

For an individual i born in village j, the issues with exclusion restriction can be clarified using the following equation:

$$y_{ij} = \tau + \theta M_i + \omega_j + \delta_{ij} \tag{4}$$

where ω_j is a village-specific effect. The key identifying assumption is that $\omega_j + \delta_{ij}$ is uncorrelated with the dialect instrument Z, the indicator that identifies whether village j has the dialect that is dominant in the province in which this village is located. Instead, if village-specific effects differ for these two types of villages (i.e., $\bar{\omega}_{Z=1} - \bar{\omega}_{Z=0} \neq 0$), the IV estimator becomes inconsistent.

The main threat to the IV strategy is that a village's dialect may be correlated with the level of health professionals, despite the former's exogeneity. For instance, individuals living in dialect-dominant villages are more likely to obtain access to better health professionals than those living outside such villages because of fewer barriers to obtaining medical services or benefits. This assumption can be tested by estimating the correlation between the dialect indicator and village-level medical benefits. Columns 1 and 2 of Table 3 reveal the results of regressing two village-level variables on the dialect indicator and the time effect. As the dataset used in this study had no direct village-level information, I calculated the logarithm of the average household medical reimbursement and (medical) insurance annuity using the sample for each village as a proxy for that village's health professionals. As depicted in Table 3, the dialect is associated with neither household medical reimbursement nor (medical) insurance. This result confirms that villages speaking the dominant dialect do not offer better health services than those without the dominant dialect. Furthermore, this suggests that people may not obtain more health benefits simply by migrating to regions with fewer dialectal barriers. This is true because, in China, most of the health service institutions (e.g., hospitals) have volunteers to help those experiencing difficulties in accessing health services.

Another challenge is that a village's dialect may be correlated with its level of economic development. For example, the nondominant dialect areas may be culturally distinct or geographically remote regions and, therefore, less likely to attract governmental support. The economic policies promoting local development tend to prioritize the needs of villages where people speak the dominant dialect of their own province; moreover, these regions are more likely to have better economic conditions. This also suggests that people in these areas are more likely to migrate to other provinces due to poor economic conditions, rather than dialect proximity. To rule out this concern, I further regressed two village-level variables—the average household assets and living cost using the sample for each village as a proxy for the its wealth—on the dialect indicator and the time effect. Table S3 reveals that the dialect is not associated with assets or the living cost per household.

In addition, I also verified whether the main results change by using another IV. In the migration literature, the migration network and history (at the village or household level) has long been used as an IV for migration. The central idea is that the migration network has a positive effect on the opportunity to migrate but no additional effects on income, education, or nutrition (Chiodi et al., 2012). Following this idea, I used the number of migrants (migration network) in the sample village as an IV for migrants. Since the migration network is a time-variant variable, I combined the IV strategy with the fixed-effects model. The main results are shown in Table S4 in the Appendix. Column 1 shows that the migration network is positively associated with individual migration, and Column 2 shows that the coefficient of interprovincial migration remains negative with a larger magnitude compared to the results using dialect proximity as an IV. This provides additional evidence on the negative impact of interprovincial migration on the health status.

TABLE 3 Relationship between village-level medical access and dialect

	(1) Medical_reimbursement	(2) Insurance_annuity
IV	−0.223	0.147
	(0.19)	(0.27)
Year	0.036	0.495**
	(0.13)	(0.19)
Constant	−64.241	−988.353**
	(265.87)	(387.90)
N	278	146
F	0.726	3.423

Note: ***1%, **5%, and *10%; standard errors are provided in parentheses. The dependent variables are the logarithm of average household medical reimbursement (in Column 1) and medical insurance annuity in the village (in Column 2). IV refers to dialect proximity.

Abbreviation: IV, instrumental variable.

5 | INTERPRETATION OF RESULTS

5.1 | Potential mechanisms

As discussed in Section 4, interprovincial migration worsens migrants' health status. In this section, I clarify some of the channels that facilitate this effect. As mentioned in Section 1, income is a pivotal mechanism associated with interprovincial migration and may positively affect the health status. The changes in daily diet, such as the changes in the consumption of non-staple food, may be positively associated with migrants' health status. However, there are two possible mechanisms through which interprovincial migration may have negative effects on health status. Migrants have a high possibility of being employed in heavy industries (e.g., mining, manufacturing, and construction) characterized by hazardous working environments, long working hours, and high work intensity, all of which likely worsen their health status. In addition to the physical health issues caused by the working environment, family separation may trigger mental health problems, including depressive symptoms and insomnia, in migrants.

Based on the ideas proposed by Baron and Kenny (1986) and Judd and Kenny (1981), I used a mediating effect model to identify the relevant mechanisms:

$$\mathrm{Y} = \mathrm{c}M_{migrant} + e_1 \tag{5}$$

$$M_{mediator} = \mathrm{a}M_{migrants} + e_2 \tag{6}$$

$$\mathrm{Y} = c'M_{migrant} + bM_{mediator} + e_3 \tag{7}$$

where $M_{migrant}$ denotes the interprovincial migration dummy, and $M_{mediator}$ represents the four mechanism variables. In Equation (5), the health status is a function of the interprovincial migration dummy, excluding the mediators. This corresponds to the results depicted in Column 3 of Table 2 for the case that uses dialect proximity as the IV and includes other control variables. As noted in Table 2, a direct negative effect is identified in Equation (5). Subsequently, Equation (6) regresses the mediators on the interprovincial migration dummy, and Equation (7) represents a function of the health status based on both the interprovincial migration dummy and mediators.

Table 4 presents the results of Equation (6). Column 1 reveals that interprovincial migration is associated with incomes that are 0.121 log points[10] higher than non-migration incomes. However, this association does not imply causality. Column 2 depicts how interprovincial migration affects the dietary pattern, which is proxied by the ratio of non-staple to staple food consumption. Dietary patterns do not appear to be significantly associated with interprovincial migration decisions. Columns 3 and 4 reveal the association between interprovincial migration, employment, and family separation using logit models. As expected, the individuals migrating to other provinces (long-distance migrants) are most likely employed in industries characterized by hazardous working environments and live away from their families.

These results suggest that changes in dietary patterns are not a mechanism through which interprovincial migration affects migrants' health status. One reason is that the increase in income is not spent on improving the quality of the migrants' daily diet. This is consistent with the finding by Cao et al. (2017) that the consumption structure of migrant worker households is dominated by survival consumption. The significance of the other three mechanisms were further verified using Equation (7). The results for the regression of the health status of migrants on the migration dummy and each of the corresponding mediators are depicted in Table 5. Column 1 reveals that migrants' health status is positively associated with the household income per capita, and the negative impact of interprovincial migration decreases (compared to the values shown in Table 2) when income is controlled for. This suggests that the increase in income is one potential mechanism through which interprovincial migration affects health.

Although Column 2 of Table 5 reveals that an increase in the ratio of non-staple to staple food is positively associated with the health status, the results shown in Table 4 negate any mediating effect, since interprovincial migration is not linked to changes in dietary patterns. The results provided in Column 3 of Table 5 reveal that the absolute magnitude of the coefficient of the migration dummy decreases (compared to values shown in Table 2) with an increase in control of the industry dummy, which is negatively associated with the health status, as expected. The results depicted in Table 2 Equation (5)) and Table 4 Equation (6) indicate that the second channel identifies a high probability of being employed in heavy industries with dangerous working environments.

Column 4 of Table 5 indicates the results of family separation. Although interprovincial migration is associated with a higher probability of separation from families (as shown in Table 4), Column 4 of Table 5 indicates no negative effects

TABLE 4 Effects of interprovincial migration on different correlates of health status

	(1) Log income	(2) Food ratio	(3) Industry	(4) Separation
Inter-provincial migration	0.121***	0.031	0.538***	1.041***
	(0.01)	(0.04)	(0.05)	(0.04)
Age	0.025***	0.040**	−0.389***	0.277***
	(0.00)	(0.02)	(0.02)	(0.02)
Age squared	0.012***	0.006	0.464***	−0.332***
	(0.00)	(0.02)	(0.03)	(0.03)
Education	0.014***	0.031***	−0.031***	0.018**
	(0.00)	(0.01)	(0.01)	(0.01)
Non-farm training	0.045***	0.249***	0.054	−0.082
	(0.01)	(0.09)	(0.09)	(0.08)
Farm training	0.023	0.020	−0.476***	0.105
	(0.01)	(0.10)	(0.11)	(0.11)
Family health	0.025***	−0.011	0.057	0.119***
	(0.00)	(0.03)	(0.04)	(0.03)
Constant	7.986***	1.445***		
	(0.07)	(0.41)		
N	172,461	170,617	34,181	42,857
Adjusted *R*-squared	0.720	0.450		
F	248.463	15.725		

Note: ***1%, **5%, and *10%; standard errors are provided in parentheses. This table shows how interprovincial migration is associated with income in Column 1(logarithm of per capita household income), dietary pattern in Column 2 (the ratio of expenditure on non-staple food to expenditure on staple food), industry type in Column 3 (an indicator for being employed in heavy industries), and separation status in Column 4 (an indicator for separating from other family members). Fixed-effect models are used in Columns 1–4.

of separation on migrants' health status, suggesting that the separation effect does not significantly affect the health status. One possible explanation is that the development of communication technology and social media has transformed migrant networks and mitigated the negative emotions caused by separation, as indicated by some migration studies (Dekker & Engbersen, 2014; González & Castro, 2007).[11]

In summary, based on the aforementioned estimates, an increase in income and exposure to hazardous working environments are two important mechanisms whereby interprovincial migration affects health status in contradictory ways. However, some more points should be emphasized here.

First, the three-equation mediation analysis in IV regressions has long been adopted in the literature (Celli, 2021). Equation (6) is estimated using the fixed-effects model, which may not appropriately identify the causal relationship between interprovincial migration and the mediating factors. Therefore, the results in Table 4 must be interpreted with caution. Many studies have discussed the causal links between migration and the two mediators (including income and industry) that have been identified as the main mechanisms in this study. For instance, a rapidly growing stream of literature indicates that rural–urban migration significantly increases migrants' income (Chan, 2013; De Brauw & Rozelle, 2008; Taylor et al., 2003); some studies show that migration increases the health risks from worksite hazard exposure (Mou et al., 2013; Walsh & Walsh, 1987). This provides additional evidence on the causal link between migration and these two mediators. In Equation (7), the main focus is not on the coefficients of the mediators, but on the change in the coefficient of migration—which is instrumented by the dialect proximity—given the inclusion of the mediators in the regression.

In addition, Dippel et al. (2020) proposed a new framework to estimate the causal mediation effects in IV regressions, showing that "while both treatment and mediator can be potentially endogenous, a single instrument suffices to identify both the causal treatment and the mediation effects." To further confirm the robustness of the result that exposure to the dangerous working environment in heavy industries is the main reason for the negative impact of interprovincial migration on migrants' health status, I also adopted the framework of Dippel et al. (2020) based on three main regressions: (i) The IV regression of health status on interprovincial migration (instrumented with dialect proximity); (ii) the

TABLE 5 Mediation effects of interprovincial migration on health status

	(1) Health	(2) Health	(3) Health	(4) Health
Inter-provincial migration	−0.116**	−0.087*	−0.109**	−0.141***
	(0.05)	(0.05)	(0.05)	(0.05)
Gender	0.101***	0.098***	0.103***	0.102***
	(0.00)	(0.00)	(0.00)	(0.00)
Age	0.017***	0.016***	0.016***	0.017***
	(0.00)	(0.00)	(0.00)	(0.00)
Age squared	−0.040***	−0.039***	−0.038***	−0.039***
	(0.00)	(0.00)	(0.00)	(0.00)
Education	0.018***	0.019***	0.019***	0.019***
	(0.00)	(0.00)	(0.00)	(0.00)
Non-farm training	0.040***	0.045***	0.049***	0.048***
	(0.01)	(0.01)	(0.01)	(0.01)
Farm training	0.057***	0.063***	0.057***	0.060***
	(0.01)	(0.01)	(0.01)	(0.01)
Family health	0.573***	0.576***	0.577***	0.576***
	(0.00)	(0.00)	(0.00)	(0.00)
Log income	0.040***			
	(0.00)			
Food ratio		0.001***		
		(0.00)		
Industry			−0.034***	
			(0.00)	
Separation				0.009**
				(0.00)
Constant	1.415***	1.757***	1.786***	1.758***
	(0.03)	(0.02)	(0.02)	(0.02)
N	172,461	170,617	172,461	172,461
Adjusted *R*-squared	0.338	0.337	0.338	0.335
F	5589.811	5489.503	5589.335	5564.577

Note: ***1%,**5%, and *10%; standard errors are provided in parentheses. This table corresponds to Equation (7), showing how the coefficient of interprovincial migration changes given that the mediators are included. The dependent variables comprise SHS. Interprovincial migration (interprovincial migrants compared to intraprovincial migrants and non-migrants as one absorbed category) is instrumented by dialect proximity in Columns 1–5.

Abbreviation: SHS, self-reported health status.

IV regression of the indicator for heavy industries (mediator) on interprovincial migration (instrumented with dialect proximity); and (iii) the IV regression of health status on the mediator (instrumented with dialect proximity), controlling for the treatment variable (interprovincial migration). The results are shown in Table 6. The coefficient of total health in Row 1 is identical to that in Column 3 of Table 2, showing that interprovincial migration leads to a decrease in the SHS score by 0.138 points. The indirect effect in Row 3 shows that a decrease of 0.224 points is driven by the mediating factor of exposure to heavy industries, which accounts for 162.77% of the total effect. This result further verifies the dominant role of the industry effect in adversely affecting migrants' health status.

However, this method could be problematic because dialect proximity may not be an ideal IV in the second and third procedures of the methods proposed by Dippel et al. (2020). Therefore, the additional analysis is only suggestive and mainly meant to facilitate a comparison to the previous results regarding the dominant role of exposure to heavy industries.

TABLE 6 Linear IV mediation analysis

Health	Coef.	Std. Err.	z	$P > \|z\|$
Total effect	−0.138	0.049	−2.83	0.005
Direct effect	0.087	0.009	9.57	0.000
Indirect effect	−0.224	0.058	−3.88	0.000

Note: The dependent variable is the SHS; the treatment variable is an indicator for interprovincial migration, and the mediator variable is an indicator for heavy industries.

Abbreviations: IV, instrumental variable; SHS, self-reported health status.

To further verify the robustness of the results, I also performed a Sobel test (Sobel, 1982) on the two identified mechanisms. The results revealed that the two *p*-values of the Sobel test corresponding to the respective mechanisms were smaller than 0.05, thereby verifying the mediation effects. However, it cannot be conclusively stated that there are only two explanations; the results are suggestive of potential and not causal mechanisms. Therefore, the findings should be interpreted with caution: the negative industry effect is larger than the positive income effect, which causes an overall negative effect of interprovincial migration on the health status.

This finding can be linked to the results in Columns 5 and 6 of Table 2, which shows that, compared to intraprovincial migrants, interprovincial migrants are more likely to report a lower SHS score, whereas the health effect differences between interprovincial migrants and non-migrants are insignificant. This is because while interprovincial migrants are more likely to be employed in heavy industries, mostly in big cities (indicating a negative effect), they have an advantage in terms of obtaining higher incomes compared to non-migrants. This suggests that the positive income effect could largely offset the negative industry effect in this case. To provide additional evidence, I demonstrate the mediating effects separately. In Table S5, Columns 1 and 3 show the mediating effects of income for interprovincial migrants compared to intraprovincial migrants, whereas Columns 2 and 4 show the effects for interprovincial migrants compared to non-migrants. Interprovincial migrants are linked to higher incomes compared to non-migrants; when income is included in the health equation, the coefficient of interprovincial migration is insignificant. This indicates a complete mediating effect of income and suggests that the positive income effect is larger in the scenario of interprovincial migrants and non-migrants compared to that of interprovincial migrants and intraprovincial migrants.

5.2 | Role of gender and assimilation

To test the robustness of the findings, I examined whether the health effect differed among men and women respondents. Since men are more likely to be employed in heavy industries, I expected the negative impact of migration on health status to be higher for men than for women. This is verified by the results depicted in Columns 1 and 2 of Table 7, which reveals that the LATE (Local average treatment effect) estimator is −0.257 for men, and the coefficient on migration is insignificant for women.

To further examine how gender influences the impact of interprovincial migration on health status, I investigated whether interprovincial migration has heterogeneous effects on the type of work done by men and women. Columns 3 and 4 of Table 7 suggest that, following interprovincial (long-distance) migration, men are more likely to be employed in heavy industries; this is consistent with the results depicted in Columns 1–2. I further identified the role of gender in interprovincial migration as follows: Compared to women, men are more likely to be employed in heavy industries with hazardous working environments, and the negative effects of interprovincial migration on health status are higher for men than for women migrants. This further confirms the dominant role of the industry effect from the gender perspective.

In addition to the role of gender, I also examined the important role of social assimilation. Interprovincial migration is likely accompanied by difficulties of social assimilation due to cultural differences (e.g., dialect dissimilarity), which may further affect the role of interprovincial migration in affecting the health status. For instance, if migrants face considerable barriers that hinder them from adapting to the local identity, they may be more likely to report poorer SHS. Most studies, like Cai and Zimmermann (2020), use self-identity as a proxy for social assimilation. However, the data does not include this type of information. Instead, I use social connections within the city (the logarithm of the amount of income presented by relatives and friends living in cities) as a proxy for potential assimilation because migrants are more likely to identify themselves as local citizens when they have close connections within the city. In Column 1 of Table S6, the interaction term of migration and potential assimilation is included in the health equation and suggests that the negative impact of interprovincial migration is smaller for those with strong assimilation compared to those with weak

TABLE 7 Impacts of interprovincial migration on men and women respondents

	(1) Health-women	(2) Health-men	(3) Industry-women	(4) Industry-men
Migration	0.047	−0.257***	0.051***	0.055***
	(0.09)	(0.06)	(0.01)	(0.01)
Age	0.016***	0.018***	−0.045***	−0.023***
	(0.00)	(0.00)	(0.00)	(0.00)
Age squared	−0.040***	−0.039***	0.054***	0.025***
	(0.00)	(0.00)	(0.00)	(0.00)
Education	0.021***	0.014***	−0.003**	−0.002
	(0.00)	(0.00)	(0.00)	(0.00)
Non-farm training	0.031**	0.059***	−0.002	0.006
	(0.01)	(0.01)	(0.02)	(0.01)
Farm training	0.104***	0.031***	−0.045***	−0.035***
	(0.01)	(0.01)	(0.02)	(0.01)
Family health	0.621***	0.536***	0.009**	0.001
	(0.01)	(0.01)	(0.00)	(0.00)
Constant	1.564***	2.063***	1.159***	0.930***
	(0.03)	(0.03)	(0.07)	(0.06)
N	84,822	87,639	84,822	87,639
Adjusted *R*-squared	0.367	0.286	0.707	0.750
F	4559.448	2596.551	37.678	18.487

Note: ***1%,**5%, and *10%; standard errors are provided in parentheses. Columns 1–2 show how the impact of interprovincial migration on the SHS differs by gender. Interprovincial migration (interprovincial migrants compared to intraprovincial migrants and non-migrants as one absorbed category) is instrumented by dialect proximity in Columns 1–2. Fixed-effect models are used in Columns 3–4, where the dependent variables are an indicator for being employed in heavy industries in Columns 1–2. Columns 1 and 3 shows the results for women and Columns 2 and 4 shows the results for men.

Abbreviation: SHS, self-reported health status.

assimilation. The potential reason is that those with strong assimilation are less likely to work in heavy industries as they may get more job information via their city connections. The results in Table S6 confirm that this is the case.

6 | ROBUSTNESS CHECKS

In this section, I present the results of several robustness checks. First, as an alternative, I divided the existing SHS into two categories as follows: I considered "Excellent" to be one category and recoded it to 1 and made "Good," "Average," "Poor," and "Very Poor" the second, non-excellent, category and recoded it to 0. Figure S3 depicts the distribution of these two categories, and Figure S4 reveals the average scores of the two categorical SHSs; the figures suggest that the scores of interprovincial migrants are slightly higher than those of their counterparts.

Subsequently, I adopted two approaches to estimate interprovincial migration's impact on the probability of a high health score being reported. First, I used a linear probability model (LPM) for estimation since the dependent variable is a binary SHS. The relevant results are presented in Column 1 of Table S7, in which the estimate using the IV-LPM that mitigates the selection bias is −0.128. This suggests that interprovincial migration decreases the likelihood of an individual reporting a high health score by 12.8 percent. Second, I used a probit model coupled with the IV strategy to estimate nonlinear treatment effects other than the aforementioned linear treatment effect. The coefficient of migration in Column 2 remains negative (−0.419), that is, interprovincial migration increases the likelihood of an individual reporting a low health score. Together, these results indicate that interprovincial migration reduces the likelihood of better health. This is consistent with the results discussed in Section 4.

Another strategy to address this issue is to replace SHS with another continuous health-related variable and re-estimate the model to test whether the negative effects persist. Hence, I used the logarithm of medical expense as a proxy for health to examine the robustness of the results. The results in Table S8 reveal that interprovincial migration is negatively associated with medical expenses. In general, health is likely to be negatively correlated with medical expenses. Healthy people may need little care, whereas sicker people need more. As such, a decline in medical expenses—especially as migrants move to a location that does not accept the NCMS insurance—suggests that the decline in health may be accompanied or even caused by gaps in the access to needed healthcare.[12] This is consistent with Hu et al. (2008), who emphasized that migrants do not qualify for some of the medical services under locality-based schemes in cities unless they pay out-of-pocket expenses.

In addition to the main dependent variable, the definition of migration may raise some concerns. The official definition of migrants in China is associated with the requirement that they live in an urban area for more than 6 months in the same year; this definition does not consider migration's impact on the individuals who move to cities for less than six months. Hence, in this study, I redefined the term migrants by reducing the period to three months and identified interprovincial migrants using this new definition. Table S9 reveals that the coefficients of interprovincial migration are positive and significant for the complete sample and for the sample of older individuals, and the magnitudes do not change significantly compared to the results depicted in Table 2.

Another threat to the validity of the main results is that interprovincial migrants may make more remittances to benefit their family members left behind in rural areas. If these remittances enhance the health of non-migrants, as a result of the aforementioned income effect, then this spillover effect will bias the impact of interprovincial migration on health. To address this issue, I analyzed a new sample that was obtained after excluding the non-migrants who received remittances from migrants. Table S10 presents the results obtained using this new sample. Columns 1 and 2 of Table S10 show the OLS and FE estimates, respectively, and columns 3 and 4 show the corresponding estimates that address the selection bias issue using IVs. As it turns out, the results are almost identical to those in Table 2, which reinforces the previous finding that interprovincial migration negatively affects the health status.

Finally, another potential issue of concern was the attrition of individuals from the survey. Following Antman (2011), I compared the impact of interprovincial migration for attritors (observed in less than 4 years) and non-attritors (observed in 4 years). The results in Table S11 show the negative effects on SHS for both attritors (Column 1) and non-attritors (Column 2), although the magnitude for attritors is much larger. Since the RCRE survey has existed since the occurrence of the first wave in 1986, the results should be interpreted with caution. Nonetheless, this analysis provides additional support for the notion that the influence of attrition is not very large.

7 | DISCUSSION AND CONCLUSIONS

Extensive literature suggests that most migrants in China find rural–urban migration a stressful process and that this process has potentially negative effects on the health status of the family members who are left behind. However, only a few studies examine the effects of this internal migration—particularly interprovincial migration, which is the preferred choice of a large number of people—on migrants' health status. In addition, to clarify relevant causal effects, one must address the issue of selection bias. The current study supports the healthy interprovincial migrant hypothesis—the positive selection of individuals to perform interprovincial migration—and overcomes selection bias by adopting the IV strategy using dialect proximity as the instrument variable, which is found to be valid for identification.

By analyzing longitudinal panel data from an RCRE survey, I reached three major conclusions that contribute an additional layer of complexity to the vast literature that links health and migration, human capital, and income. First, the results revealed that interprovincial migration negatively affects the health status of migrants in China. The estimated LATE of health status was −0.138, which suggests that interprovincial migration reduced the SHS score by 0.138 (corresponding to a 3% decrease). Compared to intraprovincial migrants, interprovincial migrants were found more likely to report a lower SHS score by 0.271, whereas the health effect between interprovincial migrants and non-migrants was insignificant.

The second conclusion pertains to two mechanisms that play a role in explaining the causal impact of interprovincial migration on health. On the one hand, interprovincial migrants' health may be improved as a result of high incomes; on the other hand, their health may be worsened by prolonged exposure to hazardous working environments. However, it seems that interprovincial migration does not affect health through family separation and changes in dietary patterns. This finding is also linked to the conclusion pertaining to the role of gender: Compared to women, men are more likely to be involved in heavy industries characterized by hazardous working environments, which suggests higher negative effects of interprovincial migration on men, rather than women, respondents' health status. The negative impact of interprovincial migration was also smaller for those with strong assimilation compared to those with weak assimilation. This may be because those with strong assimilation are less likely to work in heavy industries owing to better job information via city connections.

Finally, the main findings of this study were reinforced by conducting a series of robustness checks. However, this study was subject to some limitations: First, the main health measure used in the dataset is not sufficiently comprehensive (self-reported) and may incorporate respondents' subjective judgments. The use of other health status indicators can further enrich the results. Second, this study considered data for only four years as the main variable of interest (SHS) was not available for earlier years (e.g., surveys conducted in the 1980 and 1990s); furthermore, the data for some specific years could not be accessed, making it difficult to empirically examine the long-term impacts of interprovincial migration on migrants' health status. More studies using data for longer periods are required to investigate the presence of heterogeneities across years, for instance, to verify whether the negative health impact of exposure to hazardous environments increases with an increase in the duration of migrants' stay.

The results of this study have policy implications for the Chinese government's efforts to mitigate the negative impact of rural–urban (interprovincial) migration on migrants' health status, particularly those working in hazardous environments. Further, the government should implement measures to improve the conditions of migrants' working environments and assist the migrants in obtaining similar access to healthcare services as their urban counterparts.

Recently, China expanded its health insurance coverage in both rural and urban areas. The NRCMS (*Xin Nong He*) was piloted in the late 1990s and formally introduced in 2003. It is designed to cover all rural residents on a household enrollment basis and has proven effective in improving the healthcare access of rural residents (Sun et al., 2009). However, existing guidelines stipulate that only rural *hukou* holders can participate in local NRCMS initiatives. Meanwhile, urban areas are witnessing the evolution of a medical insurance system—the Urban Resident Basic Medical Insurance. However, to date, there is no comprehensive or universal medical coverage program targeting migrants at the national level; even at the local level, urban health insurance systems seldom incorporate all migrants free of restrictions pertaining to age, employment, or *hukou* status (Mou et al., 2013).

Notably, both the NRCMS and urban health insurance systems lack a national transfer mechanism whereby migrants, specifically interprovincial migrants, can transfer their health insurance between rural and urban areas. This can be attributed to the difference in local governments' capacity to support and subsidize medical insurance. Despite establishing national policies calling for the flexible transfer of migrants' health insurance, China still has to expend much effort to understand the health status of its substantial migrant subpopulation and find ways to improve its health by implementing effective initiatives and solutions that avert workplace injuries and provide migrants better access to healthcare services.

ACKNOWLEDGMENTS

This work is a revised version of the third chapter of my Ph.D. thesis. I would like to express my sincere gratitude to Professor Jane Golley, my primary supervisor when I did my Ph.D. at The Australian National University (ANU). She has provided continuous encouragement and excellent assistance for me to aspire towards a standard of writing and thinking which I would have otherwise considered to be out of reach. Her profound knowledge, thought-provoking advice and patient copy editing have helped to make it possible for me to complete this work. My thanks also go to other members in my supervisory panel including Professor Bruce Chapman, Dr. Kailing Shen and Dr. Nicholas Loubere. I am also deeply grateful to all the participants in seminars wherever the previous version of this work was presented. This work is supported by the National Natural Science Foundation of China (NSFC) (NO. 72003170), the National Social Science Fund of China (No. 21&ZD091), Zhejiang University-IFPRI Center for International Development Studies (No. 126000-541902/006). This work mainly use data from a survey conducted by the Research Centre for Rural Economy in China.

CONFLICT OF INTEREST

The author has nothing to disclose.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Xinjie Shi https://orcid.org/0000-0003-3965-3060

ENDNOTES

[1] See http://www.stats.gov.cn/tjsj/zxfb/202004/t20200430_1742724.html.

[2] See studies by Antman (2010), Antman (2013a, 2013b), and Ao et al. (2016) on the impact of adult child migration on the health of elderly parents who are left behind and those by Ming-Hsuan (2011), Li et al. (2015), and Gao et al. (2010) on the impact of migration on the health of left-behind children.

[3] In an additional analysis, I compare interprovincial migration with non-migrants and intraprovincial migration separately.

[4] See http://www.stats.gov.cn/tjsj/zxfb/202004/t20200430_1742724.html.

[5] I could not estimate the impact of gender because it is time invariant in this case. However, it is not the coefficient of interest here.

[6] See http://www.stats.gov.cn/tjsj/zxfb/202004/t20200430_1742724.html.

[7] Some health-related variables, such as medical expense, are used as alternatives to SHS in robustness checks.

[8] In contrast, Table S1 depicts the results based on the health equation lacking the interprovincial migration dummy. Almost all the other coefficients underwent only slight changes. This implies that these coefficients are robust to the inclusion and control for migration.

[9] The potential reason for this is explained in Section 5.

[10] In a log-linear model, the log point change is approximately 0.121 (12.1%), whereas the exact percentage point change is exp(0.121)−1 = 0.129 (12.9%).

[11] I do not have adequate data to test this explanation and, in any case, it is not the main focus of this paper.

[12] The data cannot test this and this is beyond the scope of our paper.

REFERENCES

Antman, F. M. (2010). Adult child migration and the health of elderly parents left behind in Mexico. *The American Economic Review*, *100*(2), 205–208.

Antman, F. M. (2011). The intergenerational effects of paternal migration on schooling and work: What can we learn from children's time allocations? *Journal of Development Economics*, *96*(2), 200–208.

Antman, F. M. (2013a). *How does international migration affect the health of elderly parents left behind?* Evidence from Mexico: Working Paper.

Antman, F. M. (2013b). The impact of migration on family left behind. In *International handbook on the economics of migration* (p. 293).

Ao, X., Jiang, D., & Zhao, Z. (2016). The impact of rural–urban migration on the health of the left-behind parents. *China Economic Review*, *37*, 126–139.

Baker, M., Stabile, M., & Deri, C. (2004). What do self-reported, objective, measures of health measure? *Journal of Human Resources*, *39*(4), 1067–1093.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bhugra, D., & Jones, P. (2001). Migration and mental illness. *Advances in Psychiatric Treatment*, *7*(3), 216–222.

Biao, X. (2007). How far are the left-behind left behind? A preliminary study in rural China. *Population, Space and Place*, *13*(3), 179–191.

Black, D. A., Sanders, S. G., Taylor, E. J., & Taylor, L. J. (2015). The impact of the Great migration on mortality of african Americans: Evidence from the deep south. *The American Economic Review*, *105*(2), 477–503.

Cai, S., & Zimmermann, K. F. (2020). *Social assimilation and labor market outcomes of migrants in China*. Available at SSRN: https://ssrn.com/abstract=3753895

Cao, G., Li, K., Wang, R., & Liu, T. (2017). Consumption structure of migrant worker families in China. *China and World Economy*, *25*(4), 1–21.

Celli, V. (2021). Causal mediation analysis in economics: Objectives, assumptions, models. *Journal of Economic Surveys*, 1–21. https://doi.org/10.1111/joes.12452

Chan, E. Y. Y., & Griffiths, S. M. (2010). The epidemiology of mine accidents in China. *The Lancet*, *376*(9741), 575–577.

Chan, K. W. (2013). *China: Internal migration*. The Encyclopedia of Global Human Migration.

Chen, J. (2011). Internal migration and health: Re-examining the healthy migrant phenomenon in China. *Social Science & Medicine*, *72*(8), 1294–1301.

Chiodi, V., Jaimovich, E., & Montes-Rojas, G. (2012). Migration, remittances and capital accumulation: Evidence from rural Mexico. *Journal of Development Studies*, *48*(8), 1139–1155.

Currie, J., & Stabile, M. (2003). Socioeconomic status and child health: Why is the relationship stronger for older children? *American Economic Review*, *93*(5), 1813–1823.

Cutler, D. M., Richardson, E., Keeler, T. E., & Staiger, D. (1997). Measuring the health of the US population. *Brookings Papers on Economic Activity. Microeconomics*, *1997*, 217–282.

De Brauw, A., Huang, J., Rozelle, S., Zhang, L., & Zhang, Y. (2002). The evolution of China's rural labor markets during the reforms. *Journal of Comparative Economics*, *30*(2), 329–353.

De Brauw, A., & Rozelle, S. (2008). Migration and household investment in rural China. *China Economic Review*, *19*(2), 320–335.
Dekker, R., & Engbersen, G. (2014). How social media transform migrant networks and facilitate migration. *Global Networks*, *14*(4), 401–418.
Dippel, C., Ferrara, A., & Heblich, S. (2020). Causal mediation analysis in instrumental-variables regressions. *The Stata Journal*, *20*(3), 613–626.
Fan, C. C. (2005). Modeling interprovincial migration in China, 1985-2000. *Eurasian Geography and Economics*, *46*(3), 165–184.
Frankel, J. A., & Rose, A. K. (2005). Is trade good or bad for the environment? Sorting out the causality. *The Review of Economics and Statistics*, *87*(1), 85–91.
Gao, Y., Li, L. P., Kim, J. H., Congdon, N., Lau, J., & Griffiths, S. (2010). The impact of parental migration on health status and health behaviours among left behind adolescent school children in China. *BMC Public Health*, *10*.
González, V. M., & Castro, L. A. (2007). Keeping strong connections to the homeland via web-based tools: The case of Mexican migrant communities in the United States. *Journal of Community Informatics*, *3*(3), 1–27.
Gries, T., Kraft, M., & Simon, M. (2016). Explaining inter-provincial migration in China. *Papers in Regional Science*, *95*(4), 709–731.
Gui, Y., Berry, J. W., & Zheng, Y. (2012). Migrant worker acculturation in China. *International Journal of Intercultural Relations*, *36*(4), 598–610.
Gupta, S., Pattillo, C. A., & Wagh, S. (2009). Effect of remittances on poverty and financial development in Sub-Saharan Africa. *World Development*, *37*(1), 104–115.
Hertzman, C., Power, C., Matthews, S., & Manor, O. (2001). Using an interactive framework of society and lifecourse to explain self-rated health in early adulthood. *Social Science & Medicine*, *53*(12), 1575–1585.
Howell, A. (2017). Impacts of migration and remittances on ethnic income inequality in rural China. *World Development*, *94*, 200–211.
Hu, X., Cook, S., & Salazar, M. A. (2008). Internal migration and health in China. *The Lancet*, *372*(9651), 1717–1719.
Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*, 21–37.
Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602–619.
Kaplan, R. M., & Anderson, J. P. (1988). A general health policy model: Update and applications. *Health Services Research*, *23*(2), 203.
Knight, J., & Song, L. (2003). Chinese peasant choices: Migration, rural industry or farming. *Oxford Development Studies*, *31*(2), 123–148.
Knight, J., Song, L., & Huaibin, J. (1999). Chinese rural migrants in urban enterprises: Three perspectives. *Journal of Development Studies*, *35*(3), 73–104.
Kuhn, R., Rahman, O., & Menken, J. (2006). Survey measures of health: How well do self-reported and observed indicators measure health and predict mortality. In *Aging in sub-Saharan Africa: Recommendations for furthering research* (pp. 314–342).
Li, Q., Liu, G., & Zang, W. (2015). The health of left-behind children in rural China. *China Economic Review*, *36*, 367–376.
Li, X., Stanton, B., Fang, X., Xiong, Q., Yu, S., Lin, D., Hong, Y., Zhang, L., Chen, X., & Wang, B. (2009). Mental health symptoms among rural-to-urban migrants in China: A comparison with their urban and rural counterparts. *World Health & Population*, *11*(1), 24–38.
Li, X., Zhang, L., Fang, X., Xiong, Q., Chen, X., Lin, D., Mathur, A., & Stanton, B. (2007). Stigmatization experienced by rural-to-urban migrant workers in China: Findings from a qualitative study. *World Health & Population*, *9*(4), 29–43.
Liu, Y., & Xu, X. (2016). *Language as identity: Dialectal distance and labor migration in China*. https://economics.harvard.edu/files/economics/files/liu-yuyun_language_as_identity_dialetical-distance-and-labor-migration_2-3-16.pdf?m=1454343700
Lu, S., Huang, C. C., & Rios, J. (2017). Mindfulness and academic performance: An example of migrant children in China. *Children and Youth Services Review*, *82*, 53–59.
Lu, Y. (2010). Rural-urban migration and health: Evidence from longitudinal data in Indonesia. *Social Science & Medicine*, *70*(3), 412–419.
Ming-Hsuan, L. (2011). Migration and children's welfare in China: The schooling and health of children left behind. *The Journal of Developing Areas*, *44*(2), 165–182.
Mou, J., Griffiths, S. M., Fong, H., & Dawes, M. G. (2013). Health of China's rural-urban migrants and their families: A review of literature from 2000 to 2012. *British Medical Bulletin*, *106*, 19–43.
Musgrove, P. (1993). Investing in health: The 1993 world development report of the world bank. *Bulletin of the Pan American Health Organization (PAHO)*, *27*(3), 1993.
Poncet, S. (2006). Provincial migration dynamics in China: Borders, costs and economic motivations. *Regional Science and Urban Economics*, *36*(3), 385–398.
Power, C., Manor, O., & Fox, J. (1991). *Health and class: The early years*. Chapman & Hall.
Pyakuryal, A., Tausig, M., Subedi, S., & Subedi, J. (2011). Strangers in a familiar land: The psychological consequences of internal migration in a developing country. *Stress and Health*, *27*(3), e199–e208.
Qiu, P., Caine, E., Yang, Y., Chen, Q., Li, J., & Ma, X. (2011). Depression and associated factors in internal migrant workers in China. *Journal of Affective Disorders*, *134*(1–3), 198–207.
Scharf, M. M., & Rahut, D. B. (2014). Nonfarm employment and rural welfare: Evidence from the Himalayas. *American Journal of Agricultural Economics*, *96*(4), 1183–1197.
Shen, J. (2013). Increasing internal migration in China from 1985 to 2005: Institutional versus economic drivers. *Habitat International*, *39*, 1–7.
Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology* (Vol. 13, pp. 290–312). American Sociological Association.
Song, Y., & Sun, W. (2016). Health consequences of rural-to-urban migration: Evidence from panel data in China. *Health Economics*, *25*(10), 1252–1267.
Su, Y., Tesfazion, P., & Zhao, Z. (2018). Where are the migrants from? Inter-vs. Intra-provincial rural-urban migration in China. *China Economic Review*, *47*, 142–155.

Sun, W., Wang, X., & Bai, C.-E. (2014). Income inequality and mobility of rural households in China from 2003 to 2006. *China Agricultural Economic Review*, *6*(1), 73–91.

Sun, X., Jackson, S., Carmichael, G., & Sleigh, A. C. (2009). Catastrophic medical payment and financial protection in rural China: Evidence from the new cooperative medical scheme in Shandong province. *Health Economics*, *18*(1), 103–119.

Taylor, J. E., Rozelle, S., & de Brauw, A. (2003). Migration and incomes in source communities: A new economics of migration perspective from China. *Economic Development and Cultural Change*, *52*(1), 75–101.

Walsh, A., & Walsh, P. A. (1987). Social support, assimilation, and biological effective blood-pressure levels. *International Migration Review*, *21*(3), 577–591.

Wong, D. F., He, X., Leung, G., Lau, Y., & Chang, Y. (2008). Mental health of migrant workers in China: Prevalence and correlates. *Social Psychiatry and Psychiatric Epidemiology*, *43*(6), 483–489.

Wu, R., Yang, D., Zhang, L., & Huo, J. (2018). Spatio-temporal patterns and determinants of inter-provincial migration in China 1995–2015. *Sustainability*, *10*(11), 3899.

Wurm, S. A., Li, R., & Baumann, T. (1987). *Language atlas of China*. Australian Acad. of the Humanities; Longman Group (Far East).

Zhang, J., Li, X., Fang, X., & Xiong, Q. (2009). Discrimination experience and quality of life among rural-to-urban migrants in China: The mediation effect of expectation–reality discrepancy. *Quality of Life Research*, *18*(3), 291–300.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Shi, X. (2021). Moving out but not for the better: Health consequences of interprovincial rural-urban migration in China. *Health Economics*, 1–19. https://doi.org/10.1002/hec.4468